Metabarcoding and Metagenomics 7: 1–17
DOI 10.3897/mbmg.7.95650

Research Article

Non-destructive insect metabarcoding as a surveillance tool for the Australian grains industry: a first trial for the iMapPESTS smart trap

Francesco Martoni[1], **Reannon L. Smith**[1], **Alexander M. Piper**[1], **Narelle Nancarrow**[2], **Mohammad Aftab**[2], **Piotr Trebicki**[2,3], **Rohan B. E. Kimber**[4], **Brendan C. Rodoni**[1,5], **Mark J. Blacket**[1]

1 *Agriculture Victoria, AgriBio Centre, Bundoora, VIC 3083, Australia*
2 *Agriculture Victoria, Grains Innovation Park, Horsham, VIC 3400, Australia*
3 *School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC 3052, Australia*
4 *South Australian Research and Development Institute, Plant Research Centre, Urrbrae, SA 5064, Australia*
5 *School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia*

Corresponding author: Francesco Martoni (francesco.martoni@agriculture.vic.gov.au)

Academic editor: Sarah J. Bourlat | **Received** 3 October 2022 | **Accepted** 21 December 2022 | **Published** 2 February 2023

Abstract

Surveillance and long-term monitoring of insect pest populations are of paramount importance to limit dispersal and inform pest management. Molecular methods have been employed in diagnostics, surveillance and monitoring for the past few decades, often paired with more traditional techniques relying on morphological examinations. Within this context, the 'iMapPESTS: Sentinel Surveillance for Agriculture' project was conceptualised to enhance on-farm pest management decision-making via development and deployment of smart traps, able to collect insects, as well as recording associated environmental data. Here, we compared an iMapPESTS 'Sentinel' smart trap to an alternative suction trap over a 10-week period. We used a non-destructive insect metabarcoding approach complemented by insect morphological diagnostics to assess and compare aphid species presence and diversity across trap samples and time. Furthermore, we paired this with environmental data recorded throughout the sampling period. This methodology recorded a total of 497 different taxa from 70 traps over a 10-week period in the grain-growing region in western Victoria. This included not only the 14 aphid target species, but an additional 12 aphid species, including a new record for Victoria. Ultimately, with more than 450 bycatch species detected, this highlighted the value of insect metabarcoding, not only for pest surveillance, but also at a broader ecosystem level, with potential applications in integrated pest management and biocontrol.

Key Words

aphids, biosecurity, diversity, environmental data, morphology

Introduction

Insect pests pose one of the most important threats to biodiversity, both in agricultural and natural ecosystems (Deutsch et al. 2008). These pest species can reduce yields via direct damage to crops and by vectoring plant pathogens and increase input costs, forcing growers and industries to rely on pesticides, that consequently disrupt any existing integrated pest management (IPM) systems (Pimentel et al. 2000; Ragsdale et al. 2011).

The Australian grains industry is threatened by a number of non-native aphid pests. Some of these have been introduced into the country in recent years, such as the Russian wheat aphid, *Diuraphis noxia* Kurdjumov (Yazdani et al. 2018) or *Aphis lugentis* (Petit et al. 2022), requiring the development of new control measures for plant protection and biosecurity. Other species, such as the green peach aphid *Myzus persicae*, have been present in Australia for decades and have developed resistance for several chemical insecticides (de Little and Umina 2017).

Copyright *Francesco Martoni et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Currently, aphid control in Australia relies on the early detection of new species arrivals or expanding populations, so that chemical or biological control can be deployed in a cost-effective and timely manner. Traditionally, surveillance traps placed in the field collect mixed samples which are sent to diagnostic laboratories for identification of captured insects. Here, insect diagnostics largely relies on traditional morphological examination (Hodgetts et al. 2016), which often requires advanced taxonomical skills, is focused on a limited number of taxa of interest and is made even more challenging by the increasing scarcity of trained taxonomists (Paknia et al. 2015). In recent years, traditional diagnostics targeting both the aphids and their vectored plant pathogens has been supplemented by a variety of molecular techniques, including DNA barcoding (Hebert et al. 2003), qPCR, Loop-mediated isothermal amplification (LAMP; Congdon et al. (2019)) and metabarcoding (Batovska et al. 2021). These techniques allow for standardised identification of a wide range of taxa, but each differs in the volume of samples and number of target species that can be processed at one given time (Piper et al. 2019).

Indeed, current identification methods often have time- and cost-associated limitations when required to process large numbers of specimens. Manual sorting of specimens not only requires strong entomological expertise, but is also laborious and time consuming, particularly for samples of mixed species with high numbers of specimens. For these reasons, high-throughput sequencing (HTS) technologies and techniques such as metabarcoding, are being tested worldwide for biosecurity, diagnostics and pest management purposes (Batovska et al. 2018; Tedersoo et al. 2019; Hardulak et al. 2020; Trollip et al. 2021; Young et al. 2021; Lebas et al. 2022). Metabarcoding allows DNA barcode-based identification to be conducted in a parallel manner, generating a large number of individual barcode sequences in a single reaction and, therefore, enabling the simultaneous identification of individuals in large mixed samples (Piper et al. 2019). Additionally, metabarcoding is scalable, potentially enabling the processing of tens to hundreds of samples at once (Piper et al. 2019). Scalability is a key aspect of using metabarcoding as a surveillance tool, in order to keep pace with the increasing pressure of insect pests that will increase the demand for surveillance and diagnostics.

The iMapPESTS: Sentinel Surveillance for Agriculture project, started in 2017, is a national programme of research, development and extension that was designed to put actionable information about pest and pathogen populations into the hands of Australia's primary producers to enhance on-farm pest management decision-making. The aim of the project is to lay the foundations for a national cross-industry surveillance system that – through a range of surveillance and diagnostics activities – can rapidly monitor and report the presence of airborne pests and diseases affecting major agricultural sectors across the country, including grains (https://imappests.com.au/). The project focused on the development and deployment of next generation smart traps, able to collect samples of airborne insects, viruses, bacteria and fungal spores, while also recording important environmental data (e.g. temperature, humidity, wind speed, rainfall) that is linked to sampling events and used to monitor and model movements of insects across an agricultural landscape. These smart traps, named 'Sentinels', have been deployed across multiple agricultural regions in Australia to compare them, where possible, with more traditional trapping systems.

Here, we deployed an iMapPESTS Sentinel trap at a SmartFarm in the Wimmera region (Victoria, Australia), during a 10-week trial. In order to assess the Sentinel reliability, the smart trap was compared to an alternative suction trap that is routinely used to target aphid pests in the same area. Insect samples from both traps were then processed using a non-destructive insect metabarcoding technique (Martoni et al. 2022), as a means to obtain a complete dataset encompassing both insect pests and beneficial species, such as parasitoids and predators. Furthermore, to confirm the efficacy and sensitivity of insect metabarcoding when dealing with aphid pest species, all samples were examined morphologically by expert diagnosticians to compare aphid identifications between techniques.

This enabled us to: i) assess the insect composition and diversity within the target agroecosystem, ii) compare this diversity between two different suction traps, iii) compare the use of metabarcoding analysis with more traditional morphological examinations and iv) explore the value of metabarcoding and environmental data when used to observe the presence of insects across time.

Materials and methods

Sampling

The Horsham Sentinel Trial ran for 10 weeks, from 24 September to 3 December 2021. The trial took place at the Horsham SmartFarm, in the Wimmera region of Victoria, Australia. For the duration of the trial, two trapping devices were deployed in the same area, near barley (*Hordeum vulgare*) and faba bean (*Vicia faba*) crops, at a distance of ~ 40 m from each other. Both devices were insect suction traps sampling at a height of ~ 2 m (Fig. 1A). The first device was a suction trap built by Agriculture Victoria Research team members (hereafter, AVR) that has been deployed at the Horsham SmartFarm for the past five years, used to target aphids (Hemiptera: Aphididae). The second device was an iMapPESTS Sentinel model 4 (https://imappests.com.au/what-we-do/smart-surveillance/; hereafter, Sentinel) fitted with a modified Small Aerial Vortis (SAV) sampler (supplied by Agri Samplers Ltd, High Wycombe, United Kingdom). Both devices were sampling at approx. 2 m from the ground and were fitted with a 12V axial fan to actively draw air into an omnidirectional sampling port. The

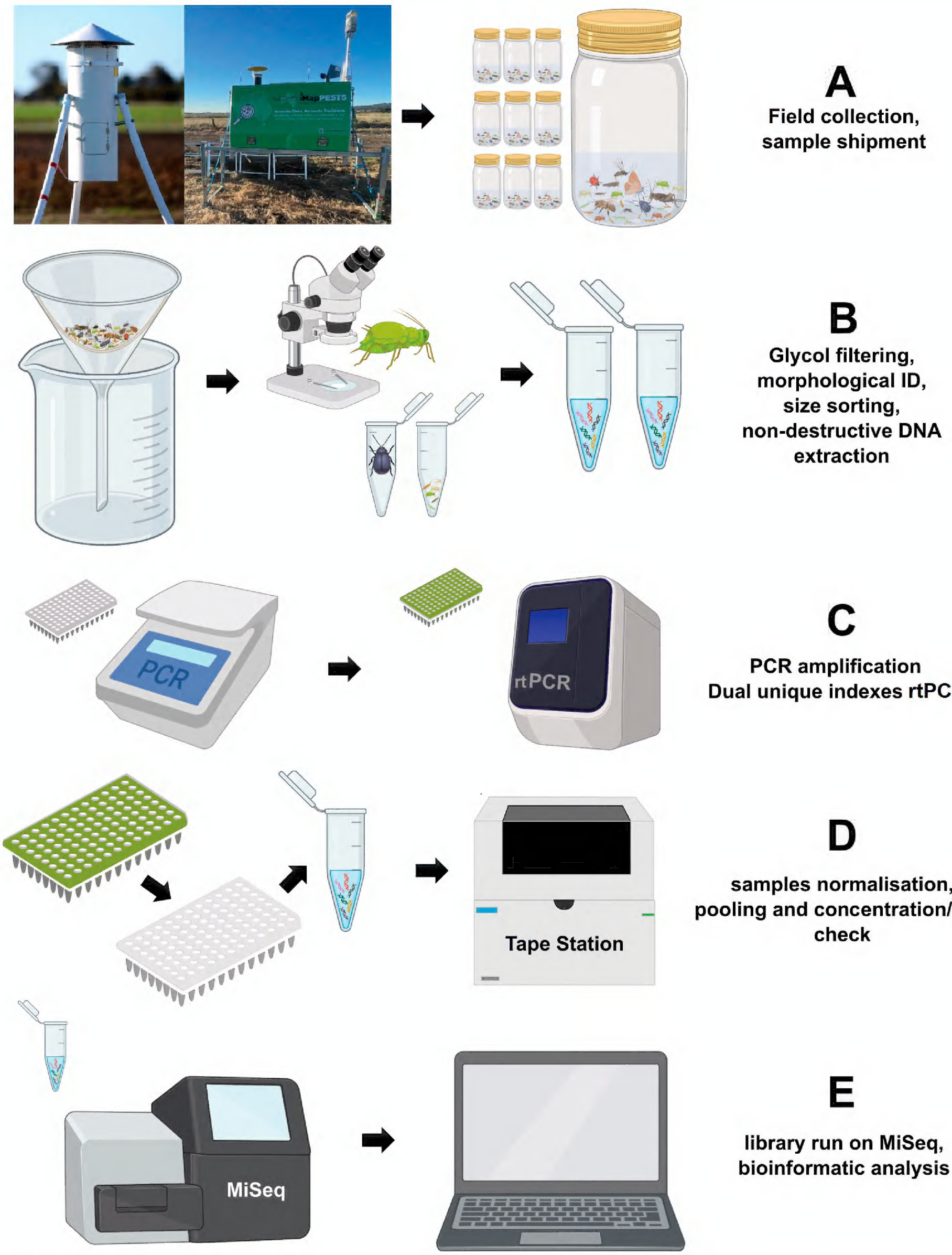

Figure 1. Non-destructive insect metabarcoding workflow. Samples are collected by the AVR trap (top left) and Sentinel trap (top right) in a propylene glycol solution (50%) (**A**); glycol is filtered and samples are examined morphologically and sorted by size, prior to non-destructive DNA extraction (**B**); partial COI barcode is amplified, Illumina adapters containing unique dual indexes are attached using real-time PCR (**C**); sample DNA concentrations are then normalised using SequalPrep normalisation plates, the library is pooled and size and concentration are inspected using a TapeStation (**D**); the final library is sequenced using an Illumina MiSeq and the data are analysed through a bioinformatic pipeline (**E**). Some details of this figure were created using BioRender (BioRender.com).

intake port and main chamber of the AVR trap were 115 mm internal diameter fitted with a rain shield with a 20 mm separation and had a sampling rate of ~ 1200 litres/min. The Sentinel SAV sampler had an intake port of 105 mm internal diameter fitted with a rain shield with a 100 mm separation and had a sampling rate of ~ 1900 litres/min. The SAV sampler design used the fan to create a vortex within an expansion chamber, to induce targets into a centrifugal spin, which directed targets into an 'off-shoot' chamber and into a sampling pot. The Sentinel 4 system was an automated platform, with a rotating carousel holding several pots integrated with the SAV sampler. The control system indexed the carousel to a new collection pot each day at midnight, for a 1-week sampling schedule before reloading the device with new pots. The AVR trap used a fan to induce a vertical flow through a mesh sieve (cloth, to collect the insect) which directed captured targets into a single sampling pot continuously until it was changed weekly. Both traps collected insects in a 50% propylene glycol solution, which was previously proved to be very effective in preserving insect DNA, as well as being easy to handle and ship due to its non-flammability (Martoni et al. 2021). While the AVR trap collected six days' worth of specimens in the same vial (which was replaced on day 7), the Sentinel trap collected the insects into daily vials, six days per week (also replaced on day 7). In total, 70 samples were collected during the 10 weeks of the trial, 10 by the AVR trap and 60 by the iMapPESTS Sentinel trap.

Sample handling and morphological identification of aphids

Samples were collected from the devices weekly by an operator (Fig. 1A). Environmental data were recorded daily by the iMapPESTS Sentinel smart trap *in loco* using an on-board automated weather station, providing precise measurements at the exact geographical location where the insects were collected. Initial morphological examination of all samples was conducted using an identification guide (Blackman and Eastop 2000) to identify and count 14 of the most important high priority target aphid species, known to be present in the region, and of interest for the grains industry (Table 1). Aphids were identified and counted, but not separated from the trap samples, and all tools and laboratory surfaces were disinfected with ethanol washes (80%–100%) between samples. After morphological examination, samples were shipped via courier to the AVR AgriBio facility for metabarcoding analysis.

Upon arrival at AgriBio, each sample was filtered to separate the insects from the collection fluid using a 0.2 mm, voile polyester fabric mesh that was previously cleaned in a bleach solution (10%) and rinsed with high grade ethanol (100%). While on the mesh, larger insects were separated from smaller specimens, stored in a different vial and then processed as separate metabarcoding samples (Fig. 1B). Each batch (14 samples) received was immediately processed for DNA extraction and PCR amplification (Fig. 1C). PCR amplicons were then stored in a -20 °C freezer until a full 96 well microtiter plate (47 samples in duplicates + one DNA extraction and one PCR negative control = 96 samples) could be simultaneously processed for library preparation. Each plate included a DNA extraction and a PCR negative control.

Table 1. Aphid species and their common names that were the focus of morphological diagnostics for both insect traps.

Species	Common name
Acyrthosiphon kondoi Shinji. 1938	Bluegreen aphid
Acyrthosiphon pisum Harris, 1776	Pea aphid
Aphis craccivora C.L.Koch, 1854	Cowpea aphid
Brachycaudus helichrysi Kaltenbach, 1843	Leaf-curling plum aphid
Brevicoryne brassicae (Linnaeus, 1758)	Cabbage aphid
Diuraphis noxia Kurdjumov, 1913	Russian wheat aphid
Dysaphis tulipae (Boyer de Fonscolombe, 1841)	Tulip aphid
Hyperomyzus lactucae (Linnaeus, 1758)	Blackcurrant-sowthistle aphid
Lipaphis pseudobrassicae Davis, 1914	Turnip aphid
Metopolophium dirhodum (Walker, 1849)	Rose-grain aphid
Myzus persicae (Sulzer, 1776)	Green peach aphid
Rhopalosiphum maidis (Fitch, 1856)	Corn aphid / Corn leaf aphid
Rhopalosiphum padi (Linnaeus, 1758)	Bird cherry-oat aphid
Uroleucon sonchi (Linnaeus, 1767)	Large sowthistle aphid

DNA extractions and library preparation

Non-destructive DNA extraction was performed using the DNeasy Blood and Tissue kit (Qiagen, Germany) with an overnight incubation period (~ 17 hours) at 56 °C as previously described (Martoni et al. 2021). The volume of ATL+proteinaseK (ratio 9:1) used depended on the number and size of insects present in each sample, ranging from 600 µL to 1 mL. A subsample of 200 µl lysate per sample was then purified on the DNeasy spin columns, following the manufacturer's instructions. After the non-destructive DNA extraction, the insects contained in each sample were preserved in high grade ethanol (100%) for further morphological examination, if required.

Polymerase chain reactions (PCRs) were performed in duplicate targeting a ~ 200 bp fragment of the standard barcode region (Hebert et al. 2003) within the *Cytochrome c Oxidase* subunit I gene (COI) using the FwhF2 – FwhR2n (Vamos et al. 2017; Marquina et al. 2018) primer pairs. Partial Illumina adapter sequences (in bold italic) were incorporated into the primers for use in 2-step PCR (FwhF2: 5'-***ACACTCTTTCCCTACACGACGCTCTTCCGATCT***GGDACWGGWTGAACWGTWTAYCCHCC-3'; FwhR2n: 5'-***GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT***GTRATWGCHCCDGCTARWACWGG-3'). The initial PCR was conducted in 25 µL reactions consisting of 14.7 µL Bovine Serum Albumin (New England Biolabs, MA, USA), 5 µL of 5x Bioline MyFi reaction buffer (Meridian Bioscience, OH, USA), 1 µL of each primer (10 µM), 0.5 µL MyFi DNA polymerase (Meridian Bioscience, OH, USA) and 2.5 µL of

template DNA. The PCR cycle started with 2 minutes at 94 °C, followed by 30 cycles of denaturation at 94 °C for 30 seconds, annealing at 49 °C for 45 seconds and extension at 72 °C for 45 seconds. Finally, the reaction ended with an extension phase of 2 minutes at 72 °C. Amplification was verified by gel electrophoresis (1% w/v agarose).

After the initial PCR amplification (Fig. 1C), the amplicons were diluted 1/10 and run through a second 6–8 cycles of real-time PCR (rtPCR) to attach unique dual indexes and the remainder of the Illumina adapter sequence (Fig. 1C). Each indexing rtPCR reaction (50 µL volume) contained 32.5 µL of ddH_2O, 10 µL of 5 x Phusion HF Buffer (Thermo Fisher Scientific, MA, USA), 1 µL of dNTP mix (10 mM), 1 µL of SYBR Green I Mix (Thermo Fisher Scientific, MA, USA) diluted 1/1000 in ddH2O, 0.5 µL Phusion DNA polymerase (Thermo Fisher Scientific, MA, USA), 4 µL of sample-specific indexing primers (2.5 µM) and 1 µL of the diluted PCR product. rtPCR cycling conditions were 98 °C for 30 s, followed by 6–8 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 30 s, with fluorescence measurement conducted in the 72 °C phase. The amplification curve was visually inspected in real time and stopped while still in the exponential phase to prevent over-amplification artefacts.

Amplicons were purified and normalised using the SequalPrep Normalization Plate Kit (Thermo Fisher Scientific, MA, USA) following the manufacturer's protocol, but eluting the final product in 15 µl instead of 20 µl (Fig. 1D). Normalised and cleaned rtPCR amplicons were then pooled together and the resulting library was quality checked, sized and quantified using a High Sensitivity D1000 ScreenTape assay performed on a 2200 TapeStation (Agilent Technologies, CA, USA) (Fig. 1D). The final pooled library was diluted to a concentration of 7 pM, spiked with 15% PhiX and sequenced using V3 chemistry (2 x 250 bp reads) across four flow cells on an Illumina MiSeq system (Illumina, CA, USA) (Fig. 1E).

Bioinformatic analysis

Bioinformatic analysis followed the pipeline generated for the iMapPESTS project and available here: https://alex-piper.github.io/iMapPESTS/local_metabarcoding.html. Raw sequence reads were demultiplexed using bcl2fastq allowing for a single mismatch in the indexes (NCBI SRA acc. number: PRJNA911921), then trimmed of PCR primer sequences using BBDuK v.38 (Bushnell et al. 2017). Sequence quality profiles were used to remove reads with more than one expected error (Edgar and Flyvbjerg 2015) or those containing ambiguous 'N' bases, then all remaining sequences were truncated to 205 bp and analysed using DADA2 v.1.16 (Callahan et al. 2016). Following denoising, amplicon sequence variants (ASVs), inferred separately from each sequencing run, were combined into a single table and chimeras were detected and removed de-novo using the *removeBimeraDenovo* function in DADA2. To further filter any non-specific amplification products and pseudogenes, the ASVs were aligned to a profile hidden Markov model (PHMM) (Eddy 1998) of the full-length COI barcode region (Piper et al. 2021) and then checked for frame shifts and stop codons that commonly indicate pseudogenes (Roe and Sperling 2007). Taxonomy was assigned using the IDTAXA algorithm of Murali et al. (2018) implemented in the DECIPHER v.2.22.0 R package, trained on an in-house COI database created for the iMapPESTS project (Piper et al. 2021), accepting only assignments with a bootstrap confidence threshold of 60% or above. To increase classification to species level, we also incorporated a BLASTn v.2.13.0 (Altschul et al. 1990) search against the same in-house database and, to reduce the risk of over-classification, we only accepted BLAST species assignments if the BLAST search agreed with IDTAXA at the Genus rank. Finally, all retained ASVs assigned to the same insect species were merged, while ASVs that could not be assigned to species, but only to a higher taxonomic rank (i.e. genus, family, order), were manually compared against the GenBank nucleotide collection (nt/nr) using the Megablast algorithm on the NCBI BLAST web server (Sayers et al. 2022). This enabled us to match or partially match the unassigned ASVs to sequences present in GenBank that may have been uploaded more recently than the reference database was created or did not pass the stringent filtering parameters defined in Piper et al. (2021). ASVs matching sequences with a similarity between 99% and 100% were labelled using the accession number of the GenBank sequence they matched (e.g. Diptera sp. XX00000). ASVs partially matching sequences, with a similarity between 96% and 98.99%, were labelled as "near" the accession number of the GenBank sequence they matched (e.g. Diptera sp. nr XX00000). Following this procedure, ASVs with a genetic similarity < 96% to any given sequence present in GenBank were manually aligned using Geneious Prime 2022.0.2 (www.geneious.com) and MEGA X (Kumar et al. 2018) and grouped into a single operational taxonomic unit (OTU) when their divergence was < 5%. Samples with less than 2000 reads, as well as ASVs with less than 0.01 relative abundance in each sample, were discarded from the dataset. Sequencing depth for all samples was assessed by generating species accumulation curves (Suppl. material 2). A heat tree was generated using the R package Metacoder v.0.3.5.1 (Foster et al. 2017), as a graphic representation of the taxonomic diversity within the dataset. After quality control, PCR replicates were merged to their original sample without any additional filtering.

Statistical analysis

For α-diversity measures, we used three complementary metrics to account for phylogenetic distance (phylogenetic diversity – pd; Faith (1992)) and abundance (Shannon diversity; Shannon (1948)), as well as simple presence-absence (Observed). The package breakaway v.4.7.9 (Willis et al. 2017) was used to predict whether species were missed due to insufficient sequencing depths. This non-linear regression model uses the abundance ratios between observed taxa within each sample to predict the

total richness (including unobserved species), following the assumption that if there were many taxa observed at very low abundance (such as with only one or two reads), there are likely many more that were observed zero times (Willis and Bunge 2015). ANOVA was then used to test whether differences in α-diversity could be explained by the trap used to collect samples. Similarly, for β-diversity analysis, we used three distance metrics (Jaccard, Aitchison and Philr) in order to consider not only presence-absence of taxa (Jaccard), but also relative abundance within a compositional data analysis framework (Aitchison metric; Aitchison et al. (2000)) and phylogenetic divergence between samples within a similar compositional framework (Philr; Silverman et al. (2017)). Principal coordinates analysis (PCoA) was used to graphically represent relationships between samples in multidimensional space using the β-diversity dissimilarity matrices. Finally, we compared β-diversity between the two trap types using permutational multivariate analysis of variance (PERMANOVA) tests using the adonis2 function from the *vegan* R package (Oksanen et al. 2020). All bioinformatic and statistical analyses were conducted within the R v.4.1.0 statistical computing environment (R Core Team 2022), using the *ALDEx2* v.1.26.0 (Fernandes et al. 2013), *phyloseq* v.1.38.0 (McMurdie and Holmes 2014), *tidyverse* v.1.3.1 (Wickham et al. 2019) packages, with figures 2–8 generated using *ggplot2* v.3.3.5 (Wickham 2016).

Results

Summary of Arthropod diversity

After quality control, 12,344,965 reads (mean 99,556 ± 60,300 per sample) were retained for subsequent analysis from a total of 124 samples. These were assigned to a total of 497 taxa, from at least 217 genera, 117 families and 15 orders, across insects and arachnids (Fig. 2). The insect order recording the most taxa were Diptera with 281 taxa (56.54%), followed by Hemiptera (53 taxa, 10.66%), Hymenoptera (44 taxa, 8.85%), Coleoptera (40 taxa, 8.05%) and Lepidoptera (28 taxa, 5.63%) (Fig. 2). The highest diversity of taxa was recorded during the final two weeks of the trial (Fig. 3).

Of the 497 taxa, only 138 (27.77%) matched the COI sequence of a barcoded described species, with an additional 39 taxa (7.85%) and 17 taxa (3.42%), respectively, matching or nearly matching a COI barcode sequence available

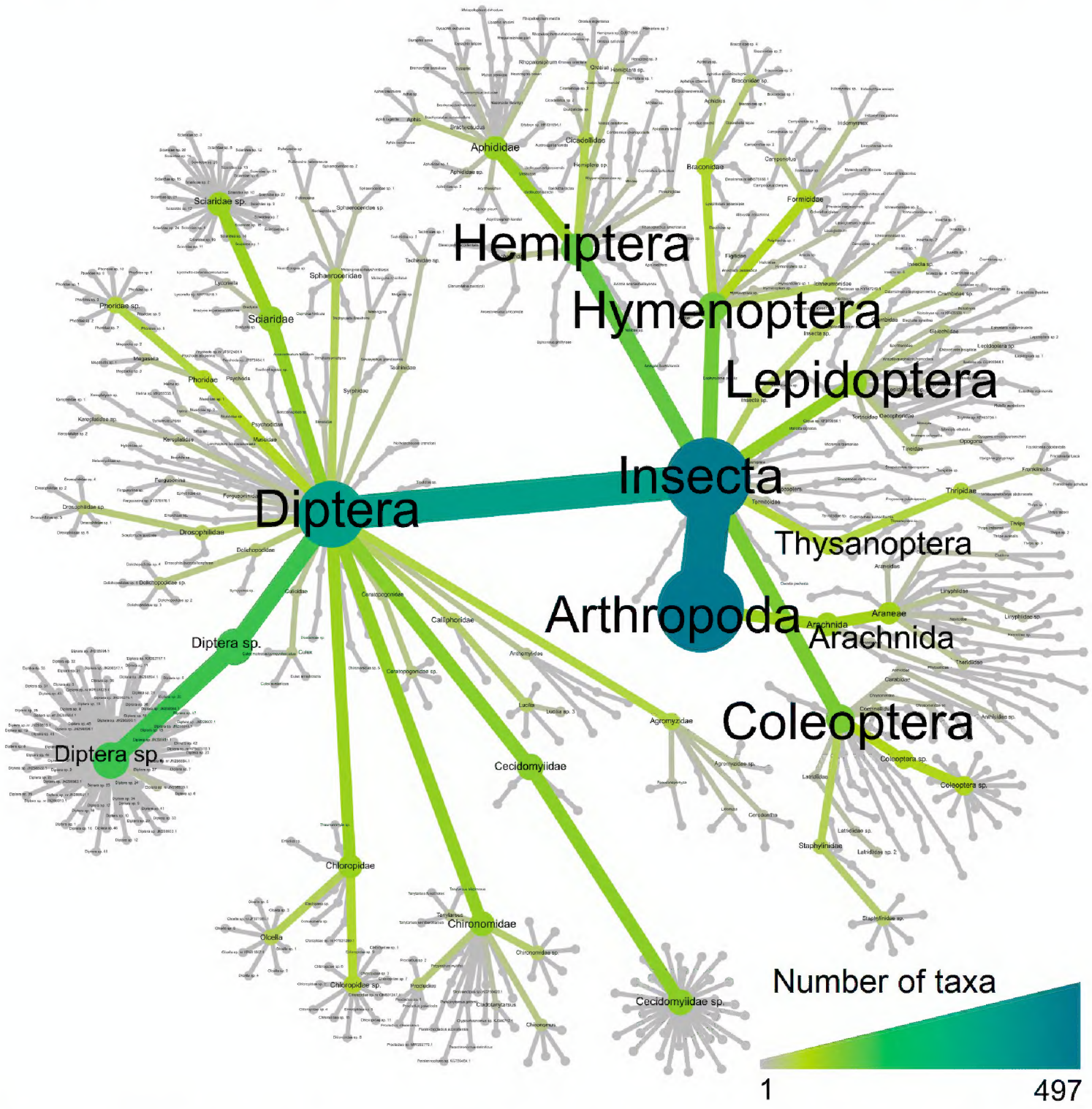

Figure 2. Heat tree summarising the relationships of the 497 taxa recorded in this study. Size and colour of branches and nodes relate to the number of taxa associated with each clade, from one (grey) to 497 (blue).

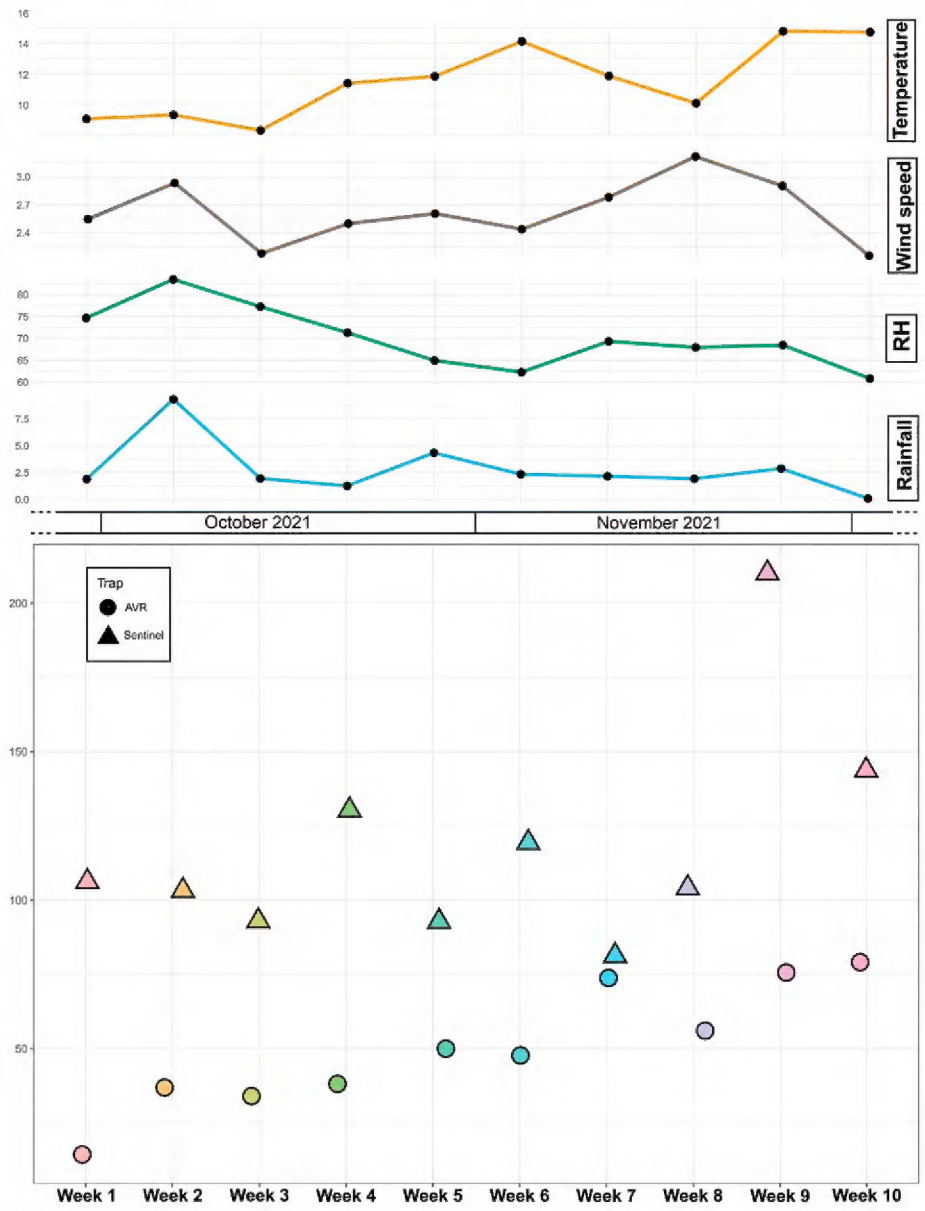

Figure 3. Environmental data recorded by the iMapPESTS Sentinel (above) associated with the observed alpha diversity across weeks (below). These graphs show the insect diversity collected by both traps (circles for the AVR trap and triangles for the Sentinel) across the 10 weeks of the Horsham trial and how it relates to rainfall, relative humidity (RH), wind speed and temperature.

in GenBank that was not identified to species level (e.g. Diptera sp. XX00000 or Diptera sp. nr XX00000) (Fig. 2, "Diptera sp."). On the other hand, 65 taxa (13.08%) could only be assigned to a genus, 157 (31.59%) only to a family, 71 (14.29%) only to an order, while 7 (1.41%) taxa could only be assigned to the class Insecta.

Within Hemiptera, 26 species of aphids were recorded, including all 14 target species (Table 1). The detected aphids included *Aphis lugentis*, only recently recorded in Australia (Petit et al. 2022), making it the first published record for the State of Victoria.

Comparison between trap types

One of the main aims of this work was to determine whether the Sentinel trap could be successfully deployed for aphid surveillance, in a similar way as the AVR trap is currently used, and what other species could be recorded in addition to the main targets. To do so, we examined species accumulation curves for all the samples analysed in this study (Suppl. material 2), to confirm that all the samples had been sequenced adequately. Additionally, when comparing the estimated diversity using the R package *breakaway* (Willis et al. 2017), this showed a variation between estimates and observed taxa that was < 1. These tests confirmed sequencing depth and taxa recovery enabled us to record most of the diversity in the community. When comparing the α-diversity between the samples collected by the AVR trap with those collected by the Sentinel trap, the latter collected significantly higher observed diversity (Observed; $F_{(1,9)} = 39.83$, $p < 0.001$) and Shannon diversity (Shannon: $F_{(1,9)} = 14.91$, $p = 0.004$), as well as collecting more phylogenetically distant taxa (PD; $F_{(1,9)} = 33.29$, $p < 0.001$) (Fig. 4), with the effect of sampling week non-significant for all comparisons ($p > 0.05$). Significant differences in community composition (β-diversity) were also found between the samples collected by the Sentinel and those collected by the AVR trap, with a separation seen between the two on PCoA plots generated using multiple distance metrics, especially Jaccard (Fig. 5). When considering just the differences in presence/absence of species (Jaccard), ADONIS tests showed that more than 10% of the variance in sample composition could be explained by the trap type ($R^2 = 0.122$, $p = 0.001$), which

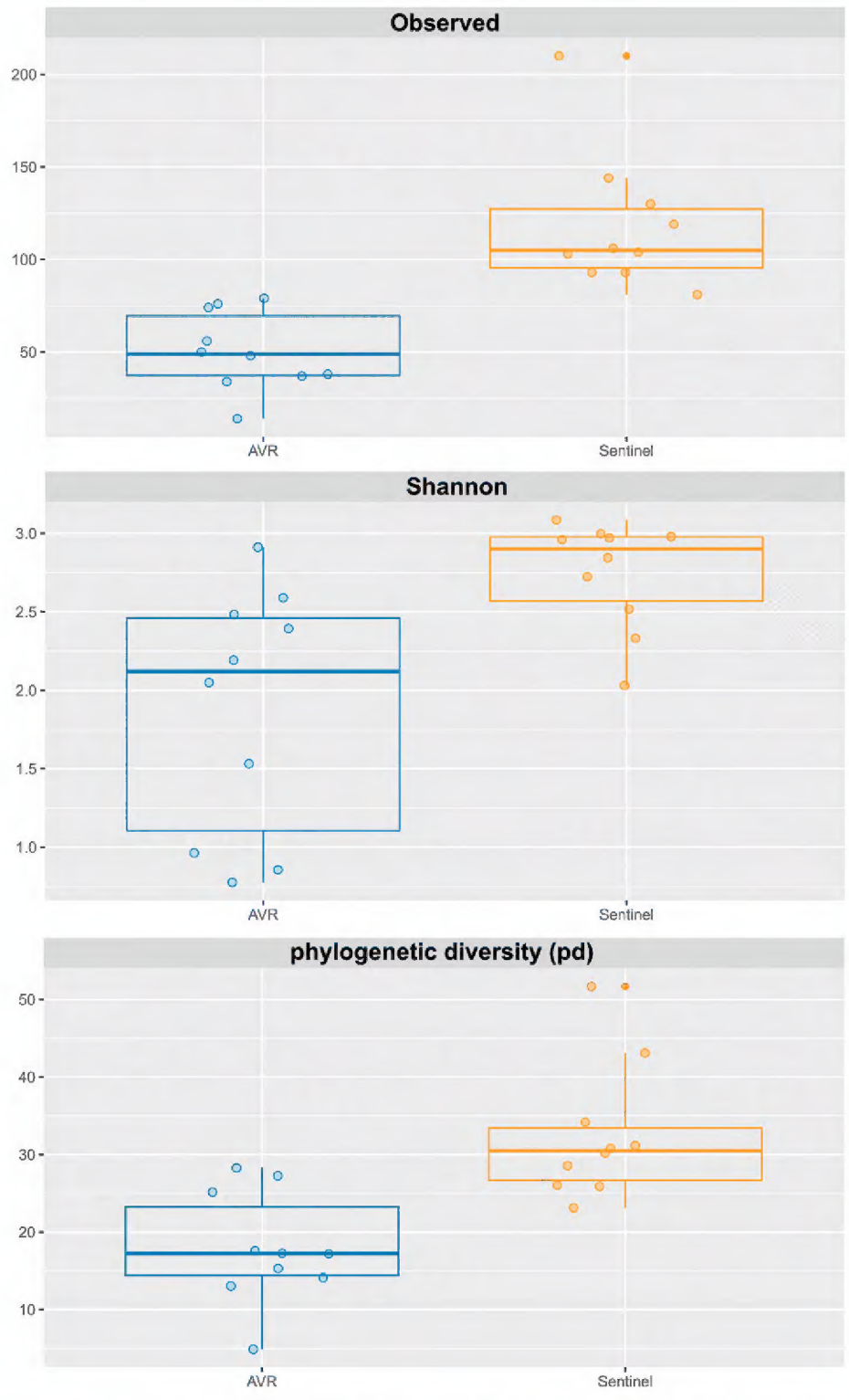

Figure 4. Box plots comparing alpha diversity between the two traps, when samples are grouped by week. Three different alpha diversity measures are shown (Observed, Shannon, phylogenetic diversity). All diversity measures show that the Sentinel recorded more diversity than the AVR trap.

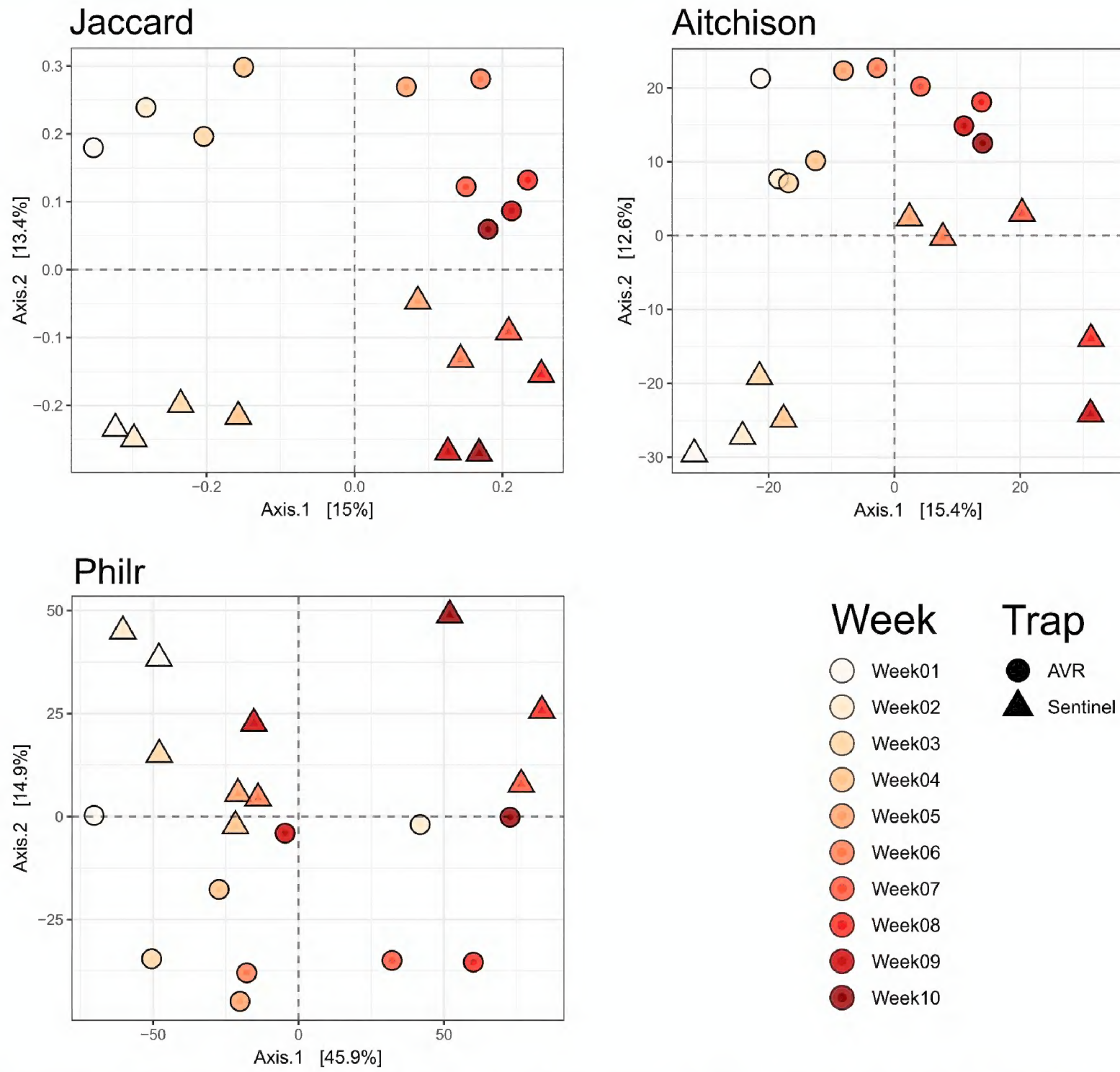

Figure 5. Principal Coordinates Analysis (PCoA) plots of distance metrics. The distance metrics used here take into account presence/absence of taxa (Jaccard), presence/absence and relative abundance (Aitchison) and presence/absence, relative abundance, and phylogenetic divergence (Philr). Samples have been merged by week, with dots representing the AVR samples and triangles representing the iMapPESTS Sentinel samples.

PERMDISP tests confirmed were due to true differences in the composition between trap types, rather than differences within group dispersion ($F_{(1,18)} = 0.66$, $p = 0.476$). When considering relative abundance as well (Aitchison), the communities caught by each trap were slightly more similar compared to presence/absence alone ($R^2 = 0.110$, $p = 0.001$); however, PERMDISP tests also found significant differences in dispersion between the two groups ($F_{(1,18)} = 31.57$, $p = 0.001$), suggesting that differences in community composition were also influenced by differences in composition within groups. This can be seen in Fig. 5, where the Sentinel samples appear to be subdivided into three separate clusters when using the Aitchison metric. Finally, when taking into account the phylogenetic relatedness between species along with their relative abundances (Philr), the communities caught by each trap appeared more similar and statistically non-significant ($R^2 = 0.107$, $p = 0.072$), with PERMDISP tests finding no significant differences in dispersion between the two traps ($F_{(1,18)} = 0.089$, $p = 0.77$).

When considering each arthropod order separately, we could determine how the two traps showed different collection patterns across the different taxa during the 10 weeks of the trial (Fig. 6). In particular, weekly detection differences between the AVR trap and the Sentinel could be observed for all orders, with the Sentinel recording taxa missed by the AVR trap (Fig. 6, in blue) and vice versa (Fig. 6, in yellow). For some orders, such as Diptera, Thysanoptera and Lepidoptera, the Sentinel trap collected taxa that were missed by the AVR trap on most occasions. For Diptera, for example, the Sentinel trap collected insects missed by the AVR trap in 549 instances (70.12%), while the AVR trap recorded insects missed by the Sentinel only 81 times (10.34%) and both trap types recorded the same taxa 153 times (19.54% – Fig. 6). This showed a ratio of about 7:1 in favour of the Sentinel trap. Other orders showed the same trend, with Hymenoptera having a ratio of 8.1:1, Lepidoptera of 4.6:1 and Thysanoptera 43:1 (Fig. 6). However, in other instances, such as for Hemiptera and Coleoptera, the two traps appeared to show more similar collection results. For Hemiptera, for example, the Sentinel trap collected insects missed by the AVR trap in 70 instances (37.63%), while the AVR trap recorded insects missed by the Sentinel 47 times

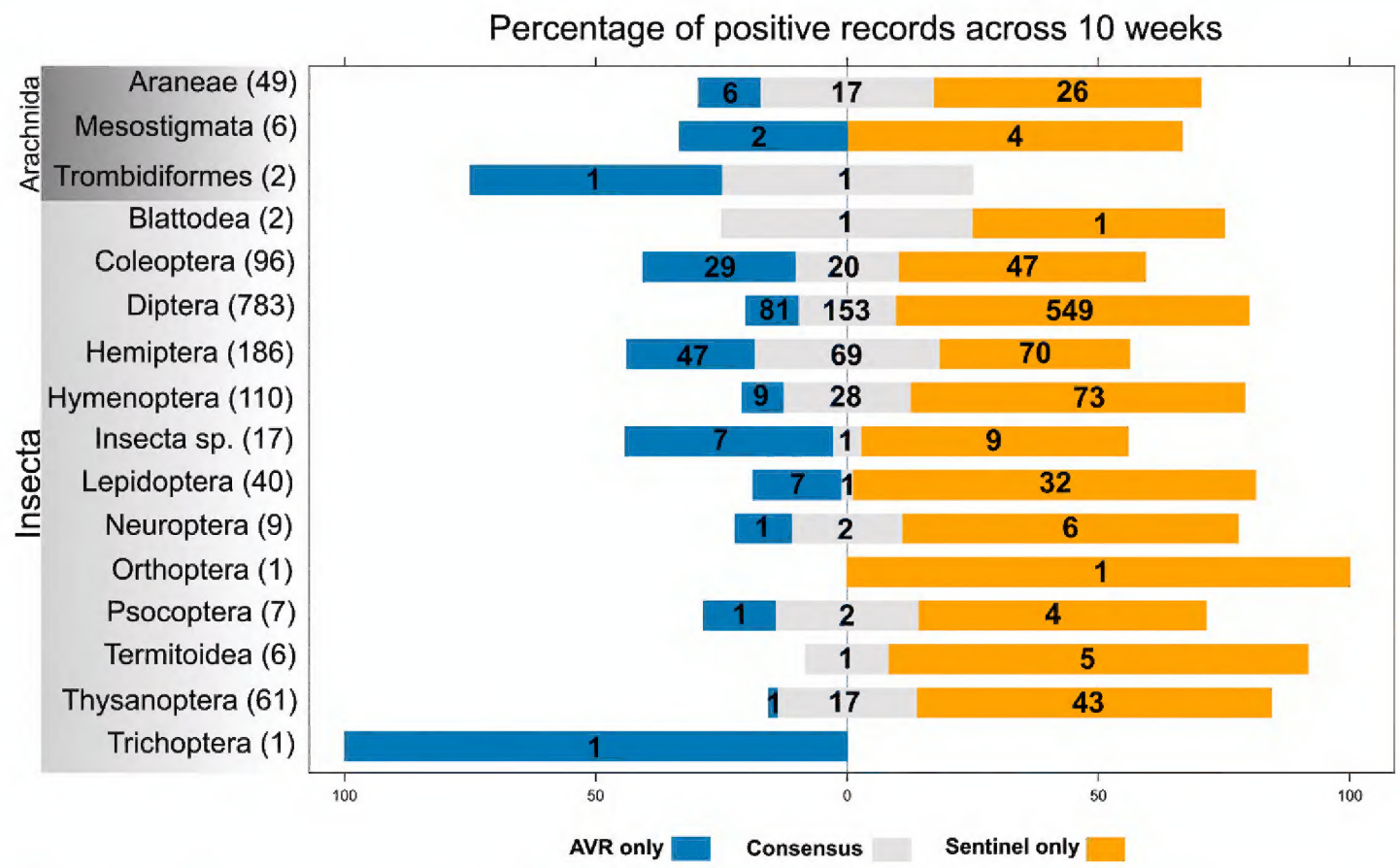

Figure 6. Percentage of weekly detections made by the AVR trap only (blue), Sentinel only (yellow) or both (consensus; grey), displayed for each taxonomic order. For each order, the overall number of records across the 10 weeks is reported in brackets. A record is intended as every instance a certain species was recorded in each week.

(25.27%) and both traps recorded the same taxa 69 times (37.1% – Fig. 6). This showed a ratio of 1.49:1 in favour of the Sentinel trap. Similarly, Coleoptera had a ratio of 1.62:1 (Fig. 6).

Morphological and molecular identification of aphids

Overall, 26 aphid species, belonging to 15 different genera, were recorded using metabarcoding (Suppl. material 1), including all the target species that were also recorded morphologically (Table 1).

Additionally, metabarcoding recorded a number of aphid taxa that did not match any COI sequence present in GenBank (Aphididae sp.1 and sp.2) and a sequence matching an undescribed *Sitobion* species (accession number MF831094).

The results of metabarcoding and morphological examination agreed for most of the samples analysed, with inconsistencies only recorded for samples with low numbers of aphids (one to six aphid individuals, Fig. 7). Metabarcoding could not record reads for one of the 14 aphid species recorded by morphological inspection for a total of 13 instances across 70 samples over the 10-week trial period, corresponding to a 90.7% congruence between metabarcoding and morphological ID. More than half of these differences (n samples = 7) occurred for the two aphid species of the genus *Rhopalosiphum*, *R. padi* (n = 2) and *R. maidis* (n = 5), with the metabarcoding not detecting one to three aphid individuals in a sample. Metabarcoding also did not detect *Dysaphis tulipae* in three samples, with six *D. tulipae* individuals not being detected in one sample. From a molecular perspective, all the aphid species recorded here had a single mismatch with the forward primer sequence, and for *Dysaphis tulipae* there was a second mismatch on the same primer which could cause further amplification bias against this species resulting in a lower representation of reads. However, no primer mismatch was present for the *Rhopalosiphum* species, for both forward and reverse primer.

In contrast, morphological examination did not record some aphid species recorded by metabarcoding in 12 instances, with four occurring for the species *Metopolophium dirhodum*, which was never recorded morphologically. These inconsistencies are difficult to ascribe to false positive results from the metabarcoding analysis, as the COI sequences obtained using metabarcoding exactly matched references sequences of these species present in GenBank and they were detected with a high number of reads (5,944 reads for *M. dirhodum*). However, due to the high sensitivity of metabarcoding, it cannot be excluded that these detections may be due to environmental contamination and/or fragments of individual aphid specimens that could not be identified using morphological examination but are present in the environment. Nonetheless, this aphid species is a common pest of grains and was previously reported from the same smart farm.

When considering single trap samples, metabarcoding sensitivity varied depending on the composition and size of samples. Metabarcoding was able to record an aphid species, based on a single aphid in samples with up to 39 aphids (sample TIP1492; Suppl. material 1), but it became more challenging with higher numbers, with single aphids at times missed in samples with 55 aphids (TIP1457; Suppl. material 1) and in samples with 57 to 74 aphids – adding to the hundreds of individuals from other species (TIP1465 and TIP1516; Suppl. material 1).

When considering the data over time (Fig. 7; Suppl. material 1), the presence of aphid species could be observed building up and/or decreasing across the trial period, depending on the different species and the environmental conditions. This result is consistent between metabarcoding and morphological examination (Fig. 7).

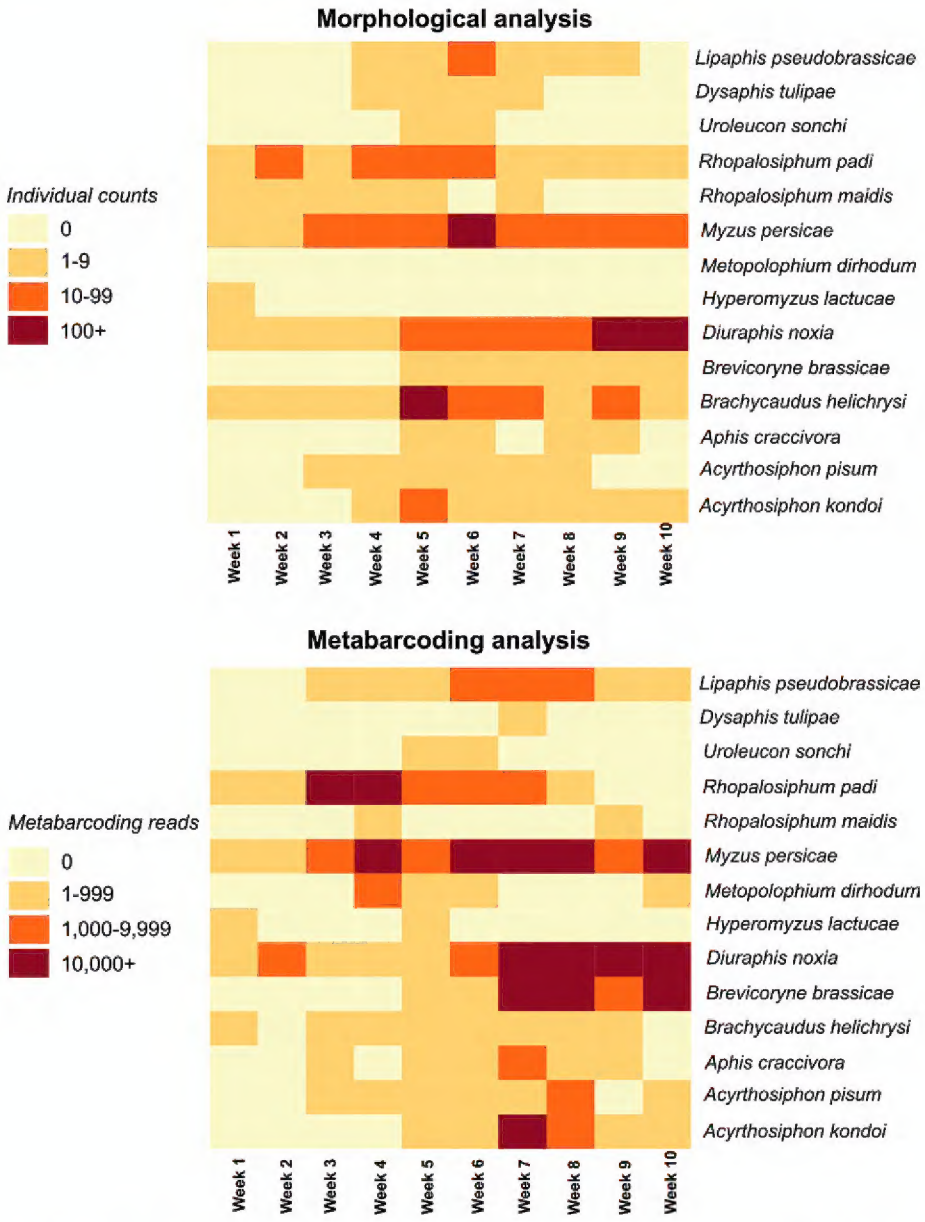

Figure 7. Heatmaps showing the variation in presence of the 14 target aphid species using morphological examination (above) and metabarcoding analysis (below), across the 10 weeks of the trial. Thresholds are colour coded.

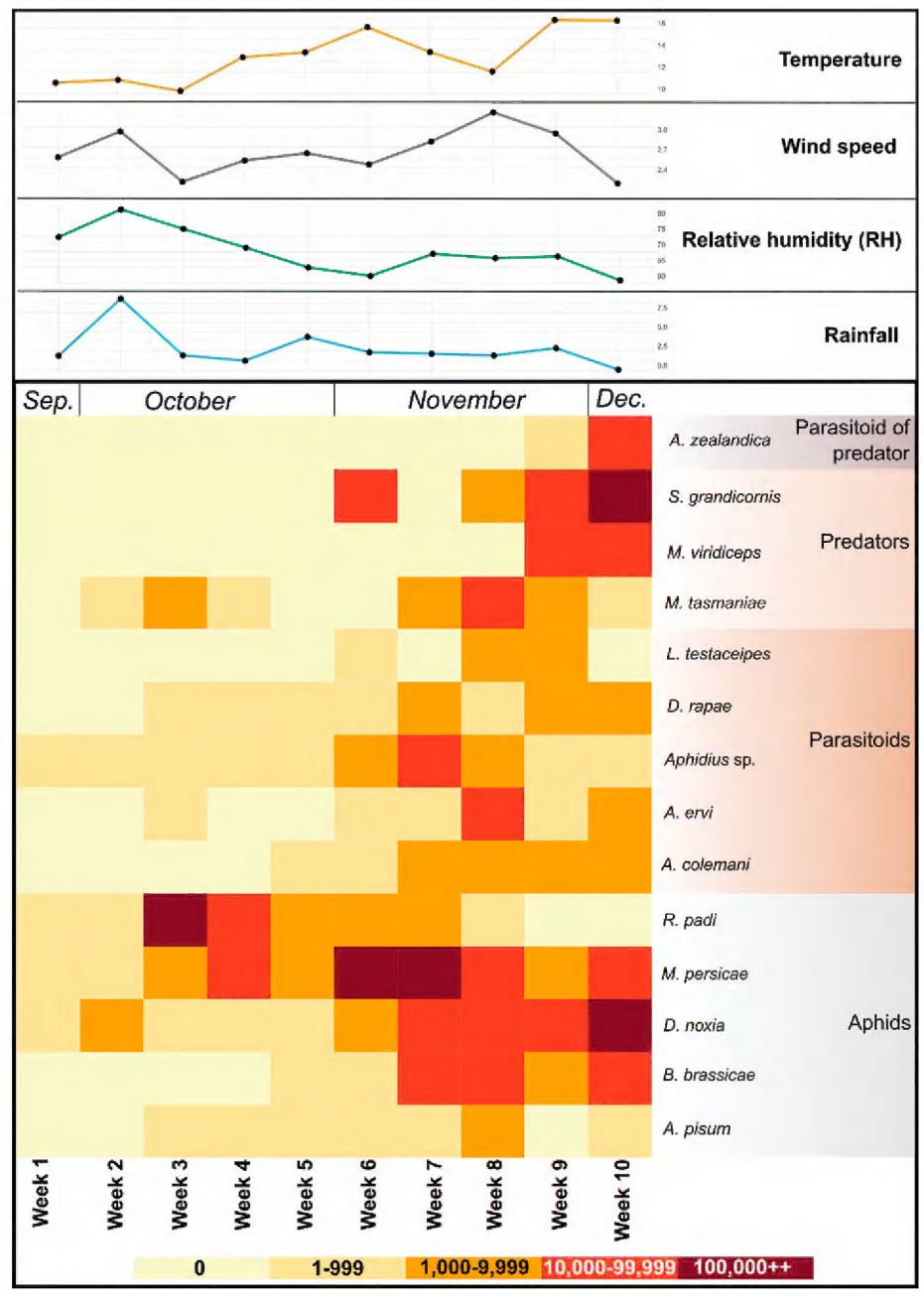

Figure 8. Environmental data (weekly averages), recorded for temperature, wind speed, relative humidity and rainfall, are reported together with the heatmap showing the variation in presence of the five most recorded target aphids together with parasitoids and predators. Thresholds are colour coded.

In general, with the progressive increase in temperatures leading to the Southern Hemisphere summer, most of the aphid species tended to increase and peak in the central weeks of the trial, then started to decrease soon after, except for the species *D. noxia*, *B. brassicae* and *M. persicae*, which peaked in the last four weeks. For example, the individual count and the number of COI reads recorded for *Diuraphis noxia* increased to a maximum peak towards the end of November (Week 9) where the temperatures were higher, with more than 10,000 reads recorded from metabarcoding (Fig. 7) and more than 400 individuals counted morphologically (Fig. 7). Similarly, an increase in *Rhopalosiphum padi* and *Myzus persicae* was recorded from week 3, peaking between weeks 4 and 6 into the trial, immediately after the peaks in rainfall and relative humidity (RH) reached on week 2. *R. padi*'s population size then started to decrease after that, as the temperatures rose (Fig. 7). *Brevicoryne brassicae's* numbers increased in the second half of the trial, while *Lipaphis pseudobrassicae*'s population started building around week 3, immediately after higher rainfalls and RH, before reaching a peak between weeks 6 and 8, then decreasing towards the end of the trial coinciding with higher temperatures.

Non-target species and environmental data across time

While the main focus of the trial was the 14 aphid species that were targeted morphologically, the use of a non-destructive insect metabarcoding technique enabled the identification of an additional 483 taxa present in the same area during the 10-week trial (Fig. 2).

Amongst these, several known beneficial insects were recorded from both the AVR and the Sentinel trap, including pollinators, parasitoids and predators. Focusing on parasitoids and predators of aphids, a number of species were recorded within the family Braconidae, across the genera *Aphidius*, *Lysiphlebus* and *Diaeretiella* (Fig. 8). These included an *Aphidius* sp. that could not be assigned to any known species, due to the lack of an identical sequence available in the reference database or in GenBank. Amongst the known predators that could be recorded were lacewings (*Micromus tasmaniae*), ladybird beetles (*Coccinella transversalis*, *Hippodamia variegate*, *Rhyzobius lophanthae*) and syrphid flies (*Simosyrphus grandicornis*, *Melangyna viridiceps*) – the larvae of which are known to predate aphids (Fig. 8; Suppl. material 1). Furthermore, we also recorded predatory insect species associated with some of the aforementioned non-target species, such as the lacewing parasitoid, *Anacharis zealandica* (Fig. 8; Suppl. material 1). Due to the large number of taxa recorded and to the fact many of these could not be matched with a known species, many additional taxa may have an ecological association with aphids.

Discussion

Insect diversity, species identification and the importance of unidentified species

The bioinformatic pipeline used here for metabarcoding was based on ASVs and enabled a more precise and unbiased species-level detection of different genetic sequences belonging to the same species when comparing these sequences against a reference database. ASVs that did not match any available record in GenBank were then grouped under an operational taxonomic unit with a 5% genetic variation, in order to not overestimate species diversity. This methodology recorded a total of 497 different taxa from 70 trap samples over a 10-week period in the grain growing Horsham region of Victoria. Of the 497 taxa detected, only 197 (39.64%) matched (or near-matched) a sequence already present in the publicly available database, such as GenBank. The remaining 300 taxa recorded here do not have an openly available COI sequence. This highlights the importance of metabarcoding studies, to explore the invertebrate diversity of regions with a scarcely documented native fauna, such as Australia.

Perfect examples of this issue are represented here by the 39 taxa recorded in this study that match sequences of unidentified insects previously uploaded in GenBank. The new records presented here can provide an important ecological tool to further understand the distribution and the role played by these taxa in an ecosystem. For example, a Phytoseiidae sp. (MF918040) and *Lycoriella* sp. (KR776019) recorded in Canada (Hebert et al. 2016), were recorded here in Victoria, Australia, suggesting the distribution of this species is not limited to North America. A *Capua* sp. and a *Scythris* sp. recorded here matched specimens (ANIC8 and ANIC 15) preserved at the Australian National Insect Collection (ANIC), that were sequenced during a "DNA Barcode Blitz" (Hebert et al. 2013), highlighting the importance of sequencing and databasing entomological collection specimens. Other unidentified species recorded here, such as an *Allodessus* sp. (KP697592) and a *Paralimnophyes* sp. (KC750464), had been previously recorded in environmental biomonitoring studies (Carew et al. 2013; Shackleton and Rees 2016), highlighting the importance of some undescribed species as potential environmental bioindicators.

Species diversity across trap types: can the broad-spectrum iMapPESTS Sentinel trap be compared to a target-specific trap?

The iMapPESTS Sentinel trap appears to be more efficient than the AVR trap since it captured more species from each insect group, such as Diptera (collecting ratio 7:1), Hymenoptera (~ 8:1), Lepidoptera (4.6:1) and Thysanoptera (43:1). This may be due to the higher sampling rate (suction "power", measured in l/min) of the Sentinel trap when compared to the AVR trap, which may affect the collection rate of larger, stronger-flying insects. This could apply to Lepidoptera and some of the larger Hymenoptera and Diptera (stronger-flying insects); however, it would probably not explain the higher collection rate for Thysanoptera (poor fliers). Interestingly, these results make the iMapPESTS Sentinel suction trap ideal to collect not only pests, but also a broader array of airborne insect populations that include beneficial insects, such as parasitoids (Hymenoptera), pollinators (Lepidoptera and Diptera) and predators (Diptera).

From a technical perspective, a number of differences are apparent when looking at the two traps. The AVR trap is a low-cost tool that was purposely built to collect aphids in the Wimmera region and has been successfully applied for the last five years. On the other hand, the Sentinel trap was built to be used across a broad range of agricultural sectors, collecting a wider range of insects and operating in diverse environmental conditions across the country. Therefore, the underlying question was not if both traps could collect the same type of insects; instead, we set out to explore whether the Sentinel could be successfully deployed to detect aphids, in a similar way as the AVR trap does and what other species could be recorded in addition to the main targets.

The results presented here show how the iMapPESTS Sentinel trap and the AVR trap were substantially comparable when collecting hemipteran insects. Although both traps recorded hemipteran taxa that the other trap did not record (25.5% of instances for the AVR trap, 37.8% for the Sentinel), the results for this order were the closest to a 1:1 ratio (1:1.48). This suggests the traps are not collecting these targets with different efficacies; instead, it suggests that two traps are enabling a better understanding of the hemipteran insect diversity of the area when paired together. Part of the differences in the taxa recorded could be due to sampling stochasticity, especially for those instances where just a single individual (or a few) was recorded for each taxon. However, the results showed numerous instances where the same species were recorded in alternation by the AVR and Sentinel traps. The iMapPESTS Sentinel trap's efficiency in collecting a specific target group of insects (in this case aphids) was shown to be comparable to the results obtained by a trap specifically designed for the task. At the same time, however, the results obtained here suggest that having more than one trap in the same crop increases the number of insect species recorded, in this case from 25.5% to 37.8% more targets. Similarly, both traps appear to show comparable results for Coleoptera, where the results are only just more biased in favour of the iMapPESTS Sentinel. A factor to consider is also the sampling frequency, which was different for each trap, with the Sentinel collecting six samples per week with each day representing a precise 24 h sample and the AVR trap sampling in the same pot for approximately 7 days until the pot is changed. As each daily Sentinel sample was processed and sequenced separately this resulted in a higher sequencing depth for a weeks' worth of Sentinel samples compared to the AVR trap, however, this did not appear to affect the species

recovery of the samples, with all samples reaching a plateau in the species accumulation curve, as well as showing minimal difference to the breakaway estimates of total diversity. A more plausible cause for the difference in the number of taxa recovered is probably the different suction pressure of the two trapping systems, with the Sentinel having a 40% higher sampling rate and increased separation between the intake port and rain shield that enables it to collect a greater diversity of flying insects, such as Diptera. However, future research should also focus on the effect of additional biological replicates (i.e. more traps) within the same surveyed area, to assess how many traps are required for a realistic assessment of the biological populations.

Aphid identification via metabarcoding and morphological assessments

In recent years, metabarcoding has been explored for insect identification for biosecurity purposes (e.g. Batovska et al. (2018); Batovska et al. (2021)). Here, metabarcoding recorded a total of 25 aphid species, belonging to 15 different genera. These detections not only included all of the target species that were recorded morphologically, but also an additional 11 species that were not targeted or could not be recorded by morphological examination. This result highlights the higher identification power of metabarcoding when compared to more traditional techniques, which are required to be more limited in taxonomic scope due to time and economic constraints, as well as taxonomical knowledge. Metabarcoding as a diagnostic tool can identify a broader set of targets, but this can only occur when metabarcoding is paired with a curated database of DNA sequences, utilising the shared efforts of many research groups in sequencing specimens. Indeed, here metabarcoding recorded aphid taxa not matching known species in public databases, but also a taxon matching an available sequence that was not identified to species level. Ultimately, the results presented here confirm that it is only by curating a DNA database with the help of strong taxonomic expertise that we can obtain species-level metabarcoding identifications.

Furthermore, the utility of metabarcoding is not limited to a diagnostic tool for the assessment of a species presence/absence. The number of DNA reads recorded during the 10-week trial showed similar patterns to the individual aphid counts performed by diagnosticians (Fig. 7). Therefore, the data reported here show how the seasonal variations in aphid populations, which are intimately linked to environmental conditions, can be assessed using metabarcoding analysis. This result suggests that long-term monitoring conducted using insect metabarcoding could be as effective as morphological examination, while also identifying a much larger number of insect species. Despite the inconsistencies reported above, limited to samples with very small numbers of aphids, the metabarcoding DNA read counts reflect the morphological counts when considering long-term surveillance over weeks and months.

The sensitivity of metabarcoding is known to be biased by a number of factors, including DNA extraction, PCR amplification and primer design (McLaren et al. 2019; Martoni et al. 2022). Importantly, this technique is commonly considered semi-quantitative and compositional in nature, meaning that the DNA reads returned for a species are only meaningful relative to the other taxa composing the sample (Gloor et al. 2017). While this semi-quantitative data may not be exactly accurate when considering a small sample size (i.e. a single sample), when considering quantitatively the reads obtained from aphids during the 10-week trial presented here, these were observed to generally match the number of individual aphids recorded morphologically. The data provides information that could be used in the future, comparing seasonality across time and to even potentially forecast population densities associated with environmental factors, such as rainfall or increases in temperatures.

Due to the metabarcoding biases mentioned above, incongruences and inconsistencies between the morphological examination and the metabarcoding analysis results are to be expected. The fact that 10 of the 13 instances where metabarcoding missed a morphologically recorded aphid were associated with just two aphid genera, *Rhopalosiphum* and *Dysaphis* and at very low individual numbers, suggests these inconsistencies are not randomly distributed. It is possible that genus-specific or species-specific factors may lead to very low number of reads recorded for one of these three species. For example, the second mismatch reported here for *D. tulipae* on the binding site of the forward primer used in this study may be a potential cause of the very low number of reads recorded for this species. Similarly, the differences in number of reads recorded for *R. padi* and *R. maidis*, two closely related species from the same genus, may be explained by DNA extraction and/or primer bias. This has been previously demonstrated for closely related species of beetles belonging to the genus *Carpophilus* and for psyllid species of the genus *Acizzia* (Martoni et al. 2022). Finally, it could be hypothesised that a destructive DNA extraction method could obtain more DNA from the insect samples; however, the non-destructive DNA extraction utilised here has been shown to perform as well as (or even better than) destructive DNA extraction methods (Martoni et al. 2021), although this is not always the case for highly sclerotised insects (Carew et al. 2018). The composition, size and diversity of the samples analysed also play an important role in biasing the DNA reads of some species in favour of others, sometimes limiting the power of metabarcoding. Ultimately, however, apart from the species of *Rhopalosiphum* and *Dysaphis*, which may require additional optimisation, the sensitivity of metabarcoding analysis in recording an aphid species that was also reported by morphological examination was 97.9%.

When considering the instances of morphological examination failing to record specimens recorded using metabarcoding, a number of factors should be considered. Firstly, one of the advantages of metabarcoding is the ability to identify partial individuals and/or from dif-

ferent life stages (i.e. nymphs and immatures; Batovska et al. (2021)) or prey DNA from the guts of predators (Simone et al. 2022), while in many cases, morphological examination can only be conducted on intact specimens. For these reasons, it remains unclear whether the detection of aphid species in the metabarcoding, but not in morphological dataset, can be attributed to partial specimens in the samples, an error of the diagnostic team or sample contamination/index switching during the library preparation step of the metabarcoding. On the other hand, in the case of *Metopolophium dirhodum*, which was never recorded morphologically, we hypothesise this species was correctly reported by the metabarcoding analysis. First of all, since no *M. dirhodum* was recorded morphologically, the hypothesis of contamination between samples should be excluded. Since no other *Metopolophium* species was recorded in this work, also the hypothesis of a misidentified closely related sequence can be excluded and the generation of a chimeric or artefactual sequence that perfectly matches (100%) *M. dirhodum* sequences in GenBank is highly unlikely. Ultimately, a plausible explanation might be that, for the aphid DNA to be present in the sample, it may have originated from the gut contents of some of the predators recorded here. For instance, in the same samples where *M. dirhodum* was recorded, we identified Staphylinidae beetles (known to be predators of invertebrates), as well as predatory spiders that could have fed upon aphid species before being collected (Suppl. material 1).

An additional point of discussion is the record of the species *Aphis lugentis* via metabarcoding. This exotic pest species has been recorded in Australia only recently (Petit et al. 2022) and the present record reports this aphid from the State of Victoria for the first time. However, additional morphological examination of the non-destructively extracted specimens could not confirm the presence of this aphid species. Indeed, even when using a non-destructive DNA extraction method, the risk of minor damage to the insect specimen, including changes in colouration or loss of appendages, cannot be prevented. This may make species-level morphological identification of some taxa challenging, if not impossible. This example raises the question whether the record of an exotic pest species should be considered true even without the presence of a voucher specimen to confirm and validate the genetic data. This adds to a discussion that has been progressing on the use of eDNA records and their potential for biosecurity and biomonitoring (Berry et al. 2021, as in the case of some plant pathogens, such as *Phytophthora* in Australia (Burgess et al. 2021). While biosecurity records should be more stringent than biodiversity assessment records, especially due to the risks and implications for market access, non-destructive insect metabarcoding may offer better chances to link a DNA sequence to an insect voucher specimen. Even when this is not possible due to the poor conditions of the specimens, as in the case of *A. lugentis* reported here, metabarcoding can offer an invaluable first detection that should be followed up with more targeted techniques, such as, morphology, qPCRs or LAMP assays, to confirm biosecurity-relevant records. Indeed, in a biosecurity context, detection of a priority pest in an area would normally provide information for future strategies, such as increased surveillance and trapping efforts, to confirm presence and delimit the extent of the outbreak. In this context, a positive record for an exotic species does not require a voucher specimen but can start an equally important biosecurity response.

Metabarcoding and environmental data: the advantage of obtaining more than a few targets.

One of the main limitations for biosecurity and surveillance is the time and expertise required for the identification of multiple targets, especially when these vary across different insect groups. Taxonomic expertise ranging across different insect orders is limited and in increasing demand (Engel et al. 2021), extremely time-consuming and, therefore, expensive. More targeted approaches, while having the advantage of being faster and cheaper, are often limited to a single taxon (e.g. qPCR and LAMP) and multiple tests/assays would be required to target additional species, consequently erasing any economic advantage. In comparison, insect metabarcoding has the ability to record, at once, as many species as can be amplified by generic PCR primers. In the case of this study, 497 taxa were recorded from 70 insect trap samples over a 10-week period, making metabarcoding the ideal tool when targeting a wide number of insect pests, even if they are spread across different families and orders. Furthermore, the advantages presented by metabarcoding do not stop at the possibility of targeting plant pests, with metabarcoding also recording beneficial insects, such as predators, parasitoids and pollinators. Such advantages have been broadly employed when using metabarcoding as a tool for biodiversity assessments (e.g. Deiner et al. (2017); van der Heyde et al. (2020)), but the benefits of this approach are not yet fully appreciated in biosecurity and pest surveillance.

Here, we demonstrated how metabarcoding records for beneficial insects, especially predators and parasitoids of aphids, mirror the records of the pests they target. For parasitoid wasps, these records are not just limited to presence/absence but, based on the number of COI reads recorded, appear to show variation in population size that is comparable to that of the aphids. These results present a potentially invaluable tool to explore the ecological network of relationships occurring amongst pests, parasitoids and predators. For example, the data reported here show how some of the parasitoids have been recorded at higher read numbers towards the end of the trial, when the pest populations were locally well-established. Some of the parasitoid species started appearing a couple weeks after the aphid populations peaked, while others were recorded during the whole trial, together with the aphids. Further trials could test whether the release of some of these parasitoid species at an earlier date could prevent the pest population to reach the same size.

Another advantage of using metabarcoding for species identification was presented by the only *Aphidius* species that could not be identified to species. Despite not being able to attribute a species-level identity due to its not matching any available DNA sequence, this *Aphidius* sp. could be recorded as a separate taxon and it was the only *Aphidius* to be recorded already from the first week of the trial. With this record, metabarcoding has provided a first insight into this species ecology, showing it appears prior to other parasitoid species of the same genus. Therefore, identifying this species and studying its biology could provide important information on its target species, potentially suggesting its use as a biocontrol agent for integrated pest management.

Finally, to fully appreciate the power of metabarcoding in unravelling ecological connections, the record of the lacewing *Micromus tasmaniae*, a predator of aphids, peaking on week 8, was followed by the record of the parasitoid wasp *Anacharis zealandica* (Figitidae) in weeks 9 and 10, a known parasitoid of lacewings. Therefore, not only could metabarcoding reveal the presence of predators of the pests, but also the presence of a parasitoid of the predator.

Conclusion

The results presented here highlight the importance of metabarcoding studies, not only as a tool for surveillance and agriculture, but also to explore the invertebrate diversity of regions with a scarcely documented native fauna.

The quantitative analysis of the reads obtained from aphids during the 10-week trial presented here, was shown to generally match the number of individual aphids recorded morphologically. This can allow comparing seasonality across time and to even potentially forecast population densities associated with environmental factors, such as rainfall or increases in temperatures. Furthermore, the detection of an exotic pest here (*Aphis lugentis*) demonstrated the strength of metabarcoding for surveillance and trapping efforts, to confirm presence and delimit the extent of the newly introduced exotic pests.

Additionally, we demonstrated how metabarcoding records for beneficial insects, especially predators and parasitoids of aphids, mirror the records of the pests they target. For parasitoid wasps, for example, these records are not just limited to presence/absence, but based on the number of COI reads recorded, appear to show variation in population size that is comparable to that of the aphids. Insect metabarcoding analysis may thus prove a useful tool for both pest surveillance and integrated pest management (IPM), with potential for monitoring populations of pests and beneficial insects simultaneously and through time, although this is still impacted by turn-around times, which may still impede timely management decisions. Ultimately, we are only beginning to scratch the surface of what may be revealed by a temporal series of metabarcoding data, such as that we have generated here. We have explored only a few of the examples that could be highlighted from the almost 500 taxa recorded here, suggesting insect metabarcoding has the potential to be used as a very important tool for IPM. This information can be used by researchers and growers to better understand the diversity of natural enemies present in an area, to provide information about whether chemical control should be used and the potential risks to established biological control agents. Additionally, paired with environmental data, metabarcoding results may enable a better understanding of how different insect populations react to environmental changes, potentially enabling forecasts of pest and beneficial insect abundance under current or future climate scenarios.

Finally, when comparing the iMapPESTS Sentinel trap to the AVR suction trap, the first appears to be more efficient than the latter in collecting a wide range of insect species. However, the two traps also appear to be substantially comparable when collecting hemipteran insects, the group for which the AVR trap was purposely built. Ultimately, this highlights the importance of understanding the biases inherent to different trap designs, especially when these biases can lead to qualitative and quantitative differences in trap catches.

Acknowledgements

The authors would like to thank MBMG editor Sarah Bourlat and reviewers Wiebke Sickel and Melissa Carew, as well as an anonymous reviewer, for the comments and suggestions contributing to improving a previous version of this manuscript. The authors would like to thank Andrew Baker (DataEffects) and Shakira Johnson (AUSVEG) for providing technical support for the iMapPESTS sentinel in the field.

References

Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2000) Logratio analysis and compositional distance. Mathematical Geology 32(3): 271–275. https://doi.org/10.1023/A:1007529726302

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3): 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Batovska J, Lynch SE, Cogan NOI, Brown K, Darbro JM, Kho EA, Blacket MJ (2018) Effective mosquito and arbovirus surveillance using metabarcoding. Molecular Ecology Resources 18(1): 32–40. https://doi.org/10.1111/1755-0998.12682

Batovska J, Piper AM, Valenzuela I, Cunningham JP, Blacket MJ (2021) Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. Scientific Reports 11(1): e7946. https://doi.org/10.1038/s41598-021-85855-6

Berry O, Jarman S, Bissett A, Hope M, Paeper C, Bessey C, Schwartz MK, Hale J, Bunce M (2021) Making environmental DNA (eDNA) biodiversity records globally accessible. Environmental DNA 3(4): 699–705. https://doi.org/10.1002/edn3.173

Blackman RL, Eastop VF (2000) Aphids on the World's Crops: an Identification and Information Guide (2nd edn.). John Wiley & Sons Ltd., Chichester, 466 pp.

Burgess TI, Edwards J, Drenth A, Massenbauer T, Cunnington J, Mostowfizadeh-Ghalamfarsa R, Dinh Q, Liew ECY, White D, Scott P, Barber PA, O'Gara E, Ciampini J, McDougall KL, Tan YP (2021) Current status of *Phytophthora* in Australia. Persoonia 47(1): 151–177. https://doi.org/10.3767/persoonia.2021.47.05

Bushnell B, Rood J, Singer E (2017) BBMerge–Accurate paired shotgun read merging via overlap. PLoS ONE 12(10): 1–15. https://doi.org/10.1371/journal.pone.0185056

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next generation sequencing: Rapid identification of macroinvertebrate bioindicator species. Frontiers in Zoology 10(1): 45–45. https://doi.org/10.1186/1742-9994-10-45

Carew ME, Coleman RA, Hoffmann AA (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? PeerJ 6: e4980. https://doi.org/10.7717/peerj.4980

Congdon B, Matson P, Begum F, Kehoe M, Coutts B (2019) Application of Loop-Mediated Isothermal Amplification in an Early Warning System for Epidemics of an Externally Sourced Plant Virus. Plants 8(5): 139–139. https://doi.org/10.3390/plants8050139

de Little SC, Umina PA (2017) Susceptibility of Australian *Myzus persicae* (Hemiptera: Aphididae) to three recently registered insecticides: spirotetramat, cyantraniliprole, and sulfoxaflor. Journal of Economic Entomology 110(4): 1764–1769. https://doi.org/10.1093/jee/tox132

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26(21): 5872–5895. https://doi.org/10.1111/mec.14350

Deutsch CA, Tewksbury JJ, Huey RB, Sheldon KS, Ghalambor CK, Haak DC, Martin PR (2008) Impacts of climate warming on terrestrial ectotherms across latitude. Proceedings of the National Academy of Sciences of the United States of America 105(18): 6668–6672. https://doi.org/10.1073/pnas.0709472105

Eddy SR (1998) Profile hidden Markov models. Bioinformatics (Oxford, England) 14(9): 755–763. https://doi.org/10.1093/bioinformatics/14.9.755

Edgar RC, Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31(21): 3476–3482. https://doi.org/10.1093/bioinformatics/btv401

Engel MS, Ceríaco LMP, Daniel GM, Dellapé PM, Löbl I, Marinov M, Reis RE, Young MT, Dubois A, Agarwal I, Lehmann PA, Alvarado M, Alvarez N, Andreone F, Araujo-Vieira K, Ascher JS, Baêta D, Baldo D, Bandeira SA, Barden P, Barrasso DA, Bendifallah L, Bockmann FA, Böhme W, Borkent A, Brandão CRF, Busack SD, Bybee SM, Channing A, Chatzimanolis S, Christenhusz MJM, Crisci JV, D'elía G, Da Costa LM, Davis SR, De Lucena CAS, Deuve T, Elizalde SF, Faivovich J, Farooq H, Ferguson AW, Gippoliti S, Gonçalves FMP, Gonzalez VH, Greenbaum E, Hinojosa-Díaz IA, Ineich I, Jiang J, Kahono S, Kury AB, Lucinda PHF, Lynch JD, Malécot V, Marques MP, Marris JWM, Mckellar RC, Mendes LF, Nihei SS, Nishikawa K, Ohler A, Orrico VGD, Ota H, Paiva J, Parrinha D, Pauwels OSG, Pereyra MO, Pestana LB, Pinheiro PDP, Prendini L, Prokop J, Rasmussen C, Rödel M-O, Rodrigues MT, Rodríguez, SM Salatnaya H, Sampaio Í, Sánchez-García A, Shebl MA, Santos BS, Solórzano-Kraemer MM, Sousa ACA, Stoev P, Teta P, Trape J-F, Dos Santos CV-D, Vasudevan K, Vink CJ, Vogel G, Wagner P, Wappler T, Ware JL, Wedmann S, Zacharie CK (2021) The taxonomic impediment: A shortage of taxonomists, not the lack of technical approaches. Zoological Journal of the Linnean Society 193(2): 381–387. https://doi.org/10.1093/zoolinnean/zlab072

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation 61(1): 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB (2013) ANOVA-Like Differential Gene Expression Analysis of Single-Organism and Meta-RNA-Seq. PLoS ONE 8(7): e67019. https://doi.org/10.1371/journal.pone.0067019

Foster ZSL, Sharpton TJ, Grünwald NJ (2017) Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. PLoS Computational Biology 13(2): e1005404. https://doi.org/10.1371/journal.pcbi.1005404

Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome datasets are compositional: And this is not optional. Frontiers in Microbiology 8: e2224. https://doi.org/10.3389/fmicb.2017.02224

Hardulak LA, Morinière J, Hausmann A, Hendrich L, Schmidt S, Doczkal D, Müller J, Hebert PDN, Haszprunar G (2020) DNA metabarcoding for biodiversity monitoring in a national park: Screening for invasive and pest species. Molecular Ecology Resources 20(6): 1542–1557. https://doi.org/10.1111/1755-0998.13212

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings. Biological Sciences 270(1512): 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert PDN, deWaard JR, Zakharov EV, Prosser SWJ, Sones JE, McKeown JTA, Mantle B, La Salle J (2013) A DNA 'Barcode Blitz': Rapid Digitization and Sequencing of a Natural History Collection. PLoS ONE 8(7): e68535. https://doi.org/10.1371/journal.pone.0068535

Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, Pedersen S, Jannetta P, deWaard JR (2016) Counting animal species with DNAbarcodes: Canadian insects. Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences 371(1702): e20150333. https://doi.org/10.1098/rstb.2015.0333

Hodgetts J, Ostojá-Starzewski JC, Prior T, Lawson R, Hall J, Boonham N (2016) DNA barcoding for biosecurity: Case studies from the UK plant protection program. Genome 59(11): 1033–1048. https://doi.org/10.1139/gen-2016-0010

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35(6): 1547–1549. https://doi.org/10.1093/molbev/msy096

Lebas B, Adams I, Al Rwahnih M, Baeyen S, Bilodeau GJ, Blouin AG, Boonham N, Candresse T, Chandelier A, De Jonghe K, Fox A, Gaafar YZA, Gentit P, Haegeman A, Ho W, Hurtado-Gonzales O, Jonkers W, Kreuze J, Kutjnak D, Landa B, Liu M, Maclot F, Malapi-Wight M, Maree HJ, Martoni F, Mehle N, Minafra A, Mollov D, Moreira A, Nakhla M, Petter F, Piper AM, Ponchart J, Rae R, Remenant B, Rivera Y, Rodoni B, Roenhorst JW, Rollin J, Saldarelli P, Santala J, Souza-Richards R, Spadaro D, Studholme DJ, Sultmanis S, van der Vlugt R, Tamisier L, Trontin C, Vazquez-Iglesias I, Vicente CSL, Vossenberg BTLH, Wetzel T, Ziebell H, Massart S (2022) Facilitating the adoption of high-throughput sequencing technologies as a plant pest diagnostic test in laboratories: A step-by-step description. Bulletin OEPP. EPPO Bulletin. European and Mediterranean Plant Protection Organisation 52(2): 394–418. https://doi.org/10.1111/epp.12863

Marquina D, Andersson AF, Ronquist F (2018) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources 19(1): 90–104. https://doi.org/10.1111/1755-0998.12942

Martoni F, Nogarotto E, Piper AM, Mann R, Valenzuela I, Eow L, Rako L, Rodoni BC, Blacket MJ (2021) Propylene glycol and non-destructive DNA extractions enable preservation and isolation of insect and hosted bacterial DNA. Agriculture 11(1): 77–77. https://doi.org/10.3390/agriculture11010077

Martoni F, Piper AM, Rodoni BC, Blacket MJ (2022) Disentangling bias for non-destructive insect metabarcoding. PeerJ 10: e12981. https://doi.org/10.7717/peerj.12981

McLaren MR, Willis AD, Callahan BJ (2019) Consistent and correctable bias in metagenomic sequencing experiments. eLife 8: e46923. https://doi.org/10.7554/eLife.46923

McMurdie PJ, Holmes S (2014) Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking. Bioinformatics (Oxford, England) 31(2): 282–283. https://doi.org/10.1093/bioinformatics/btu616

Murali A, Bhargava A, Wright ES (2018) IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences. Microbiome 6(1): 140–140. https://doi.org/10.1186/s40168-018-0521-5

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2020) vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Paknia O, Rajaei ShH, Koch A (2015) Lack of well-maintained natural history collections and taxonomists in megadiverse developing countries hampers global biodiversity exploration. Organisms, Diversity & Evolution 15(3): 619–629. https://doi.org/10.1007/s13127-015-0202-1

Petit S, Weyland JJ, Brumley C, Kehoe MA, Wang C (2022) First record of *Aphis lugentis* in Australia, tended by native ants on *Senecio odoratus*. Austral Ecology 47(1): 137–141. https://doi.org/10.1111/aec.13056

Pimentel D, Lach L, Zuniga R, Morrison D (2000) Environmental and economic costs of nonindigenous species in the United States. Bioscience 50(1): 53–65. https://doi.org/10.1641/0006-3568(2000)050[0053:EAECON]2.3.CO;2

Piper A, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience 8(8): giz092. https://doi.org/10.1093/gigascience/giz092

Piper AM, Cogan NOI, Cunningham JP, Blacket MJ (2021) Computational evaluation of DNA metabarcoding for universal diagnostics of invasive insect pests. BioRxiv 435710. https://doi.org/10.1101/2021.03.16.435710

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/

Ragsdale DW, Landis DA, Brodeur J, Heimpel GE, Desneux N (2011) Ecology and management of the soybean aphid in North America. Annual Review of Entomology 56(1): 375–399. https://doi.org/10.1146/annurev-ento-120709-144755

Roe AD, Sperling FA (2007) Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. Molecular Phylogenetics and Evolution 44(1): 325–345. https://doi.org/10.1016/j.ympev.2006.12.005

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST (2022) Database resources of the national center for biotechnology information. Nucleic Acids Research 7(50): D20–D26. https://doi.org/10.1093/nar/gkab1112

Shackleton M, Rees GN (2016) DNA barcoding Australian macroinvertebrates for monitoring programs: Benefits and current short comings. Marine and Freshwater Research 67(3): 380–390. https://doi.org/10.1071/MF14331

Shannon CE (1948) A mathematical theory of communication. The Bell System Technical Journal 27(3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Silverman JD, Washburne AD, Mukherjee S, David LA (2017) A phylogenetic transform enhances analysis of compositional microbiota data. eLife 6: e21887. https://doi.org/10.7554/eLife.21887

Simone Y, Chaves C, van der Meijden A, Egeter B (2022) Metabarcoding analysis of different portions of the digestive tract of scorpions (Scorpiones, Arachnida) following a controlled diet regime shows long prey DNA half-life. Environmental DNA 4(5): 1176–1186. https://doi.org/10.1002/edn3.311

Tedersoo L, Drenkhan R, Anslan S, Morales-Rodriguez C, Cleary M (2019) High-throughput identification and diagnostics of pathogens and pests: Overview and practical recommendations. Molecular Ecology Resources 19(1): 47–76. https://doi.org/10.1111/1755-0998.12959

Trollip C, Carnegie AJ, Dinh Q, Kaur J, Smith D, Mann R, Rodoni B, Edwards J (2021) Ophiostomatoid fungi associated with pine bark beetles and infested pines in south-eastern Australia, including *Graphilbum ipis-grandicollis* sp. nov. IMA Fungus 12(1): 24–24. https://doi.org/10.1186/s43008-021-00076-w

Vamos EE, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1(5): e14625. https://doi.org/10.3897/mbmg.1.14625

van der Heyde M, Bunce M, Wardell-Johnson G, Fernandes K, White NE, Nevill P (2020) Testing multiple substrates for terrestrial biodiversity monitoring using environmental DNA metabarcoding. Molecular Ecology Resources 20(3): 732–745. https://doi.org/10.1111/1755-0998.13148

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer, New York. https://doi.org/10.1007/978-3-319-24277-4

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. Journal of Open Source Software 4(43): e1686. https://doi.org/10.21105/joss.01686

Willis A, Bunge J (2015) Estimating diversity via frequency ratios. Biometrics 71(4): 1042–1049. https://doi.org/10.1111/biom.12332

Willis A, Bunge J, Whitman T (2017) Improved detection of changes in species richness in high diversity microbial communities. Applied Statistics 66(5): 963–977. https://doi.org/10.1111/rssc.12206

Yazdani M, Baker G, DeGraaf H, Henry K, Hill K, Kimber B, Malipatil M, Perry K, Valenzuela I, Nash MA (2018) First detection of Russian wheat aphid *Diuraphis noxia* Kurdjumov (Hemiptera: Aphididae) in Australia: a major threat to cereal production. Austral Entomology 57(4): 410–417. https://doi.org/10.1111/aen.12292

Young RG, Milián-García Y, Yu J, Bullas-Appleton E, Hanner RH (2021) Biosurveillance for invasive insect pest species using an environmental DNA metabarcoding approach and a high salt trap collection fluid. Ecology and Evolution 11(4): 1558–1569. https://doi.org/10.1002/ece3.7113

Supplementary material 1
All the taxa (ASVs) recorded in this study

Author: Francesco Martoni, Reannon L. Smith, Alexander M. Piper, Narelle Nancarrow, Mohammad Aftab, Piotr Trebicki, Rohan B. E. Kimber, Brendan C. Rodoni, Mark J. Blacket

Data type: table of taxa recorded per sample, across traps and weeks.

Explanation note: The table lists all the taxa (ASVs) recorded in this study. For each taxon, the ASV identification number, the associated COI sequence and the taxonomical identification obtained in this work are provided. Number of reads are reported for each sample and total number of reads are reported per taxon and per sample. Samples are ordered by week and subdivided by trap type (AVR or iMapPESTS sentinel).

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.95650.suppl1

Supplementary material 2
Species accumulation curve

Author: Francesco Martoni, Reannon L. Smith, Alexander M. Piper, Narelle Nancarrow, Mohammad Aftab, Piotr Trebicki, Rohan B. E. Kimber, Brendan C. Rodoni, Mark J. Blacket

Data type: figure

Explanation note: Species accumulation curve showing the sequencing depth for each sample analysed in this study. The graph shows that most samples, independently of the trapping methods, have reached a plateau, indicating that sequencing depth was not a factor influencing the diversity of taxa detected.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.95650.suppl2